DOCUMENT RESUME

ED 209 909 EC 140 629

AUTHOR	Burke, Nancy E.; And Others
TITLE	A Film Evaluation Checklist to Describe Instructionally Relevant Film Characteristics: A Rater Reliability Study.
INSTITUTION	National Technical Inst. for the Deaf, Rochester, N. Y.
SPONS AGENCY	Department of Health, Education, and Welfare, Washington, D.C.
REPORT NO	NTID-PS-42
PUB DATE	Jan 81
NOTE	74p.; For related documents, see EC 140 611-630.
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Check Lists; *Deafness; *Instructional Films; *Media Selection
IDENTIFIERS	*Film Evaluation Checklist

ABSTRACT

A Film Evaluation Checklist, developed at the National Technical Institute for the Deaf, describes instructionally relevant film characteristics. The instrument is organized in terms of structural characteristics of the medium, modes of concept emphasis using film techniques, and organization of instructional material. Raters independently rated three captioned films repeated at three weekly sessions. Interrater agreements exceeding 70% were reached for each single viewing. It was concluded that films may be reliably and accurately rated after a single viewing. The Film Evaluation Checklist should be useful for selecting films for instructional use or identifying films with certain instructionally relevant characteristics. Appendixes include a brief history of the Visual Media Research Project, a list of project consultants and reviewers, a copy of the Film Evaluation Checklist, a copy of the Master Criterion Rating Checklist, a description of films and criterion ratings assigned to films, a rater's guide to the Film Evaluation Checklist, an outline for film rater training, tables with statistical data, and a copy of the Revised Film Evaluation Checklist. (Author/SB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED209909

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

A Film Evaluation Checklist to Describe Instructionally Relevant Film Characteristics: A Rater Reliability Study

Nancy E. Burke
Fred J. Dowaliby
Barbara G. McKee

Department of Educational Research & Development
National Technical Institute for the Deaf

Paper Serier No. 42
January, 1981

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

NTID

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

EC140629

NTID's principle goal in doing research is to influence the education, training and career placement of deaf citizens through systematic examination of issues related to deafness. As one part of NTID's total research effort, the Department of Research and Development conducts descriptive and experimental research. Research findings are used in the development of programs and materials in the areas of learning and instruction, personal and social growth, and career development of deaf students. This document was developed in the course of an agreement with the U.S. Department of Health, Education and Welfare.

Table of Contents

	Page
Introduction	1
Development of Instrument	2
Method	5
Materials	5
Films	5
Rater's Guide	6
Raters	6
Criterion Ratings	7
Procedures	7
Analysis and Results	8
Question 1	8
Question 2	11
Question 3	13
Discussion	15
Dropped Items	15
Section and Item Characteristics	15
Suggestions for Revisions and Potential Users	19
Limitations and Suggestions for Future Research	20
Summary	21
Reference Notes	23
References	24
Appendices	26

Appendices

A. History of the Visual Media Research Project

B. Project Consultants and Reviewers

C. Film Evaluation Checklist

D. Master Criterion Rating Checklist

E. Description of Films and Criterion Ratings Assigned to Films

F. Rater's Guide to the Film Evaluation Checklist

G. Outline for Film Rater Training

H. Additional Data

I. Revised Film Evaluation Checklist

Abstract

A Film Evaluation Checklist to describe instructionally-relevant film characteristics was developed at NTID. The instrument is organized in terms of: structural characteristics of the medium, modes of concept emphasis using film techniques, and organization of instructional material. Raters independently rated three captioned films repeated at three weekly sessions. Interrater agreements exceeding 70 percent were reached for each film and session. It was concluded that films may be reliably and accurately rated after a single viewing. The Film Evaluation Checklist should be useful for selecting films for instructional use or identifying films with certain instructionally-relevant characteristics for research purposes.

Introduction

Films have been considered as important and useful instructional materials for many years. As a result, both teachers and educational researchers are interested in the educationally important characteristics of films. Teachers are often required to make rental or purchase decisions about films and researchers are often interested in how different film characteristics contribute to different instructional outcomes. A prerequisite for either of the above tasks is the ability to describe or rate certain characteristics of films that are thought to be important to the desired educational outcomes.

The impetus for this project was a larger study that intended to examine interactions among certain characteristics of visual media (primarily films), context variables (such as how the film was used in the classroom), student characteristics (such as cognitive style) and student learning from the films. (For a brief history of the visual media project, see Appendix A.) The authors realized there was a need to describe or categorize certain film characteristics independently of their use in the classroom. The Film Evaluation Checklist was developed to fulfill this purpose. The study reported in this paper examined the interrater reliability of the instrument and determined the number of viewings of a film necessary for acceptable consensus among raters. Specifically, the investigators were attempting to answer the following three questions:

1. Are there items on the instrument that cannot be rated with a reasonable degree of consensus even after three viewings of a film? Such items should be modified or discarded.
2. Did agreement among raters improve with increased viewings of a film? This is important because it indicates the number of viewings necessary for maximum (asymptotic) reliability.

3. Could inexperienced raters, after a 2-hour training session, use the instrument with enough precision to obtain ratings close to or to equal to those of a skilled rater who reiteratively examined the films? Did their ratings improve in the sense of approaching the skilled rater's criterion ratings after two or three viewings?

Development of the Instrument

Most attempts to develop film evaluation instruments have focused upon the organization of educational content rather than the means of presenting content to viewers. As Salomon and Snow (1968) point out, the structure of the medium must be considered as well as the content.

The evaluation of film structure may be performed at two levels of analysis: micro and macro. Microstructure consists of the basic components of film; individual shots made by the camera, cuts linking shots, and sequencing of these components (Hochberg & Brooks, 1978). Patterns of presentation, however, make up a film's macrostructure. Examples of structural components at this level include relationships among verbal, visual, and auditory presentation modes (Salomon & Snow, 1968), information density, and how instructional concepts are made salient.

Macrostructural film components are the focus of this study for three reasons. The first is that they are more easily evaluated once they are defined, and a film may be evaluated by a trained rater at a single viewing. This is clearly advantageous when a rater has a large number of films to evaluate on a routine basis.

Secondly, the most powerful instructional effects are expected with macrostructural components. For example, field dependence cognitive style (Witkin, Moore, Goodenough, & Cox, 1977) has been shown to significantly interact with relative amounts of pictorial and verbal information presented. For example, Marantz and Dowaliby (Note 1) and Koran, Snow, and MacDonald (1971) have demonstrated that the use of concrete pictorial

examples throughout instruction relatively facilitates subsequent learning performance of field-dependent learners. No difference due to pictorial examples was indicated for relatively field independent learners. This finding suggests that the presence of pictorial examples may serve as a compensatory function as instructional mediators for field dependent learners. This is particularly critical for our purposes since recent findings (Parasnis and Long, 1979) indicate that hearing-impaired learners are relatively field dependent as a group. More expansive discussions of these and related findings may be found in Witkin et al (1977) and Cronbach and Snow (1977).

The third reason for the macrostructural level of analysis is that films at NTID are captioned, providing both verbal and pictorial information in the visual mode. The film evaluation instrument, thus, must distinguish among verbal and pictorial information.

A 50-item prototype instrument was developed through the use of several sources. First, a large-scale literature review of research on film structure, aptitude-treatment interactions, and perceptual processes was conducted to identify as many instructionally-relevant film characteristics as possible. Second, 10 people in the field were interviewed to obtain their views about film characteristics that may affect student learning and student attitudes. These resource people included: researchers, NTID media specialists, classroom instructors, instructional media producers, and instructional developers (See appendix B for list of interviewees).

Initial refinement of the prototype instrument was accomplished with two pilot reliability studies in which the three investigators independently used the checklist to rate four films. Items with poor interrater agreement were dropped or rewritten, yielding the final 24-item version used in the present study.

The present version of the instrument included four parts or sections (See Appendix C for copy of instrument). These four sections were labeled:

I. Production Characteristics

II. Pictorial Characteristics

III. Verbal Characteristics

IV. Organizational Characteristics

The first section of the checklist, Production Characteristics, is for rating the formats and general purposes of the presentation. Items about whether the presentation is in a film or videotape format, in color or black and white, and captioned or not are included. Also important are instructional concerns such as the appropriate age-groups of viewers, currentness of the presentation, the instructional purpose, and the type of presentation (e.g., whether in lecture or discussion format). In general, the items in this section are those used in most media evaluation checklists.

Presentation of pictorial material is rated in the Pictorial Characteristics section of the checklist. Items focus on the use of motion photography, the kinds of artwork (e.g., cartoons, charts, graphs, conventional photography), the use of close-ups, color, and pictorial examples; and the overall adequacy of the pictorial presentation.

Ratings in the Verbal Characteristics section assess the amount and difficulty of the verbal information presented. Verbal information in our study meant information provided through captions, although, in other settings, this would mean information provided through the sound track. Since captions seldom provide a word for word duplication of the sound track, several items on the adequacy of captioning are included. Deaf postsecondary students, in general, have attained lower reading levels and are not as familiar with elaborate syntactic structures (e.g., relative clauses) and specialized vocabulary when compared to hearing counterparts (Johnson, 1976; McKee & Hausknecht, 1980; Quigley, Wilbur, Power, Montanelli & Steinkamp, 1976; Wilbur, 1977). These are important considerations when evaluating a captioned film. For example, Braverman, Harrison, Dowaliby, and Bowker (Note 2) found that students' synthesis of captioned

material was decreased when captions with greater syntactic complexity were used. Reading comprehension is also affected by the language level of captioning (Braverman, Note 3; Braverman and Hertzog, 1980). Braverman and her associate have not found changes in comprehension to be related to captioning rates between 60 and 120 words per minute. Consequently, it is important to judge the difficulty and adequacy of the captions.

The Organizational Characteristics of the instructional content and its presentation must also be evaluated. As noted above, the explicitness of the conceptual structure of the material and its content area have been shown to have powerful interactions with the field-dependence cognitive style.

The evaluation of both the structure of the material and the ways in which it is presented should provide a balanced research or evaluation instrument useful for numerous research and pedagogical applications.

Method

Materials

Films

Three films were selected for this study based on the following characteristics:

(1) similarity in length

(2) representative variation of content

(3) captioned

(4) scripts available in hard copy

(5) photographed in color

(6) available in videocassette format

The films used in the study were presented in videotape format because of the availability of a VCR and monitor in meeting facilities near the department. The three films employed in the study were:

(1) Meters, Liters, and Kilograms, (1974), 23 min.

(2) Principles of Quality Concrete, (1969), 23 min.

(3) Cross Country Ski Technique, (1973), 25 min.

See Appendix E for a brief description of the films.

Raters Guide

The raters were instructed on the use of the instrument during a 2-hour training session which included (film clip) examples of the items. Parts of a chemistry film, Elements, Compounds and Mixtures (1961), were used to supply examples. A Rater's Guide (Appendix F) and a Film Evaluation Checklist (Appendix C) were given to each rater. After the explanation of the items was completed, the raters rated a sample film, The Time of Your Life (1974). After the ratings of this practice film were completed, the ratings and items which caused difficulty were discussed (see Appendix G, Outline for Film Rater Training).

Raters

Ten faculty or staff members agreed to participate as raters in this study. One rater had to drop out after one week for reasons unrelated to the study, reducing the number of raters to nine. None of the raters had training in film production, although two had advanced training in instructional development and were familiar with instructional media. The remaining seven raters were in unrelated fields of education.

Criterion ratings

Each film and its captioning script were studied intensively by the first author[1] to establish the criterion ratings for the films. These criterion ratings were particularly important because they were the validation criteria against which the accuracy of the raters responses were judged. Quantitative definitions were developed for as many items as possible (see Appendix D, Master Criterion Rating Checklist).

For items that were inherently qualitative (e.g., items on "instructional purpose" or those on "organization"), the first author made judgments based on an intensive study of the film and its script. For example, for Section IV (organization), she constructed topic outlines to aid her in evaluating the organization of the film. Some of the qualitative guidelines are outlined in Appendix F, Rater's Guide, and others are incorporated into the checklist items (Appendix C). The actual rating given to each of the three films on each item by the first author is included in a description of the film (Appendix E).

Procedures

The raters were trained together in the two hour training session previously described. Following the training session, the raters attended rating sessions at weekly intervals to independently rate the series of three captioned films. The same three films were rated in counterbalanced order at each of the three sessions in order to minimize any fatigue or practice effect for the second and third film shown at any sessions. Since each film ran

---

[1]The first author is currently a doctoral candidate who specializes in perceptual and cognitive development, and does research on perception of visual media.

about 25 minutes and completing the checklist after each film required about 5 minutes, each rating session (including breaks between films) lasted slightly less than two hours.

Because of scheduling difficulties, three rating sessions were scheduled for each week. Films were shown in a different counterbalanced order for each session during the week and across weeks. Because several raters changed their schedules, strict within-rater counterbalancing of films did not always work. However, across weeks and raters the counterbalancing generally held true.

In order to maximize the independence of ratings, checklists were turned in after each session and raters were asked not to discuss the films among themselves. The films were presented with the sound turned off so that all verbal information corresponded with the captioning script. Each rater was given a number which only s/he knew for signing completed checklists. These numbers are not the numbers used for designating raters in the Tables.

Analyses and Results

Question 1:

Are there items on the instrument that cannot be rated with a reasonable degree of consensus even after three viewings of a film? An adjunct question was also examined: Did any of the nine raters differ enough from his/her peers to adversely influence the results?

Analysis. It was arbitrarily determined that any item that did not have perfect agreement by six of the nine raters (2/3 level of agreement) on at least three of the nine ratings would be dropped from further analysis. Since each film was rated three times, this preset cut-off would allow the raters to disagree during the first two viewings of the film, provided that six of the nine raters were in perfect agreement during the third and final viewing.

Results. Table 1 presents items that reached the preset level of agreement among raters. If two-thirds (6 of 9) of the raters achieved perfect agreement for a given film and rating session, that item was considered to have scored a "hit;" if this preset level was not reached, that item was considered to have scored a "miss." Based on the above data, three items were dropped from further analysis. These were items 17, 22 and 24.

Since there were only nine raters, the investigators were concerned that one or two people might adversely influence the results if their ratings were significantly different from the other raters. To check for this possibility, the mode for each item (by film and session) was calculated. This yielded a total of 189 times each rater could agree with his/her peers (21 items by three films by three rating sessions). The percentage of time each rater agreed with the mode was then calculated by simply counting the number of agreements and dividing by 189. The average or mean percentage of agreement could then be calculated for each person. It was decided to discard the data of any rater whose level of agreement with his/her peers was two standard deviations or more from this mean. Though two raters had overall percentage agreements with the modal ratings somewhat lower than the others, this difference was not statistically significant and their data were retained in the analysis.

Table 1

Items With 2/3 Level of Rater Agreement (6 of 9 raters) or More for Each Film at Each Session (S_1,S_2,S_3). (1 = hit, 0 = miss)

		Film								
		Meters			Concrete			Ski		
Section	Item	S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
I: Production	1.	0	0	0	1	1	1	0	0	0
	2.	1	0	0	0	0	0	1	1	1
	3.	1	1	1	1	1	1	1	1	1
	4.	1	1	1	1	1	1	1	1	1
II: Pictorial	5.	1	1	1	1	1	1	1	1	1
	6.	1	1	1	1	1	1	1	1	1
	7.	0	0	1	1	1	1	1	1	1
	8.	1	1	1	1	1	1	1	1	1
	9.	0	1	0	1	1	0	1	1	1
	10.	1	1	1	1	1	1	1	1	1
	11.	1	1	1	1	1	1	1	1	1
	12.	1	1	1	1	0	1	1	0	0
	13.	1	0	0	1	1	1	0	0	0
III: Verbal	14.	0	1	1	0	0	0	1	0	1
	15.	0	0	0	1	1	1	1	0	1
	16.	1	1	1	0	0	0	0	0	1
	(17.)[a]	(0)	(0)	(0)	(0)	(1)	(1)	(0)	(0)	(0)
	18.	1	1	0	1	1	1	1	1	1
	19.	0	1	0	1	1	1	1	1	1
	20.	1	1	1	1	1	1	1	1	1
	21.	1	1	1	0	0	1	1	0	1
IV: Organization	(22.)[a]	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
	23.	0	0	0	0	0	0	1	1	1
	(24.)[a]	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(1)	(0)
	Totals	14	15	13	16	15	16	18	14	18
	Films Totals	Meters 42 (67%)			Concrete 47 (75%)			Ski 50 (79%)		
	Sessions Totals	Session 1 48 (76%)			Session 2 44 (70%)			Session 3 47 (75%)		
	Grand Total	139 (74%)								

[a] Dropped Item

Question 2

Did agreement among raters improve with increased viewings of a film?

Analysis. In order to examine whether the raters' responses became more similar with increased viewings of a film, each rater's response on an item was compared with the mode of that item for each film and session. It was decided to use item modes as a basis of comparison rather than item means, because an item mean could obscure large differences between raters. For example, for item #18 (Does screen positioning of captions show which actor is speaking), 50% "yes's" and 50% "no's" would give an item mean of "sometimes," but such a mean would not reflect the split among the raters.

Results. Agreement with item modes are presented in Table 2. If the raters' responses became more similar with each rating session, the numbers under each film would become increasingly larger. However, an inspection of Table 2 shows that this is not the case. In fact, agreement with item modes was highly consistent across time. Interrater agreement across all ratings was 79%. Agreement with item modes across the three times was 81, 78, and 79% for the three times, respectively. These results indicate that rater agreement did not increase with repeated viewings of a film.

Table 2

Number of Ratings Matching Modal Ratings (total possible = 9 in a cell) for Each Item by Film and Session (S_1,S_2,S_3).

Section	Item	Film: Meters S_1 S_2 S_3	Concrete S_1 S_2 S_3	Ski S_1 S_2 S_3	Item Totals
I: Production	1.	4 5 5	8 8 6	8[b] 4 5	53 (65%)
	2.	6 8[b] 8[b]	8[b] 5 8[b]	9 8 9	69 (85%)
	3.	9 9 9	8 7 9	9 8 9	77 (95%)
	4.	9 8 9	8 7 9	9 9 9	76 (94%)
II: Pictorial	5.	8 8 9	9 8 9	9 9 9	78 (96%)
	6.	9 9 9	9 9 9	9 9 9	81 (100%)
	7.	4 3 5	7 8 7	9 9 9	61 (75%)
	8.	8 8 9	7 8 8	9 9 8	74 (91%)
	9.	5 6 5	6 7 5	6 7 7	54 (67%)
	10.	7 9 8	9 7 7	6 8 7	68 (84%)
	11.	9 8 8	9 8 8	9 9 7	75 (93%)
	12.	9 9 8	8 5 7	8 8 6	68 (84%)
	13.	6 5 4	6 6 7	4 4 5	47 (58%)
III: Verbal	14.	5 7 6	5 5 5	6 7 7	53 (65%)
	15.	5 5 5	7 7 8	7 5 6	55 (68%)
	16.	7 8 7	4 4 4	5 5 8	52 (64%)
	(17.)[a]	(5)(5)(5)	(5)(6)(6)	(5)(5)(5)	(47)(58%)
	18.	8 7 5	9 7 8	9 9 8	70 (86%)
	19.	5 6 5	8 8 8	8 8 8	64 (79%)
	20.	9 9 8	9 9 7	9 9 7	76 (94%)
	21.	8 7 6	4 5 6	6 5 7	54 (67%)
IV: Organization	(22)[a]	(4)(4)(4)	(4)(4)(4)	(5)(6)[b](5)	(41)(51%)
	23.	4 4 5	5 4 5	6 6 7	46 (57%)
	(24.)[a]	(8)[b](5)(5)	(3)(4)(7)	(9)[c](6)(4)	(51)(63%)
	Totals	144 148 143	153 142 149	160 155 157	1351 (79%)
	Film Totals	Meters 435 (77%)	Concrete 444 (78%)	Ski 472 (83%)	
	Session Totals	Session 1 457 (81%)	Session 2 445 (78%)	Session 3 449 (79%)	
	Grand Total	1351 (79%)			

[a] Dropped Item

[b] Bimodal Item

[c] Trimodal Item

Question 3:

Could inexperienced raters, after a 2-hour training session, use the instrument with enough precision to obtain ratings close or equal to those of a skilled rater who reiteratively examined the films? Did their ratings improve (in the sense of approaching the skilled rater's criterion ratings) after two or three viewings?

Analysis. While questions 1 and 2 were concerned with the similarity of the rater's judgments, question 3 is concerned with the accuracy of the judgments. Clearly, the "correct" response to any item on the instrument is open to debate; however, the investigators felt that the care with which the first author developed the criterion ratings assured that these ratings were as close to being "true" as possible. The criterion ratings made by the principal investigator were then compared with the modal ratings of the nine raters for each item, film, and session. For this analysis, the concept of a "hit" and "miss" was used again. A "hit" means the modal ratings of the nine raters equaled the criterion rating.

Results. The results of each item by film and rating session are shown in Table 3. The three rating sessions yielded agreements of 81, 78 and 76%, respectively, with the criterion ratings. Agreement for the three films was 75% (Meters), 78% (Concrete), and 83% (Ski). An overall level of 78% agreement with the criterion ratings was reached. There was no increase in accuracy (meaning closer agreement with criterion) for more viewings of the film. In fact, there was a slight decrease in accuracy. Hence, neither an increase in agreement (question 2) nor an increase in accuracy is obtained through increased viewings of a film.

Table 3

Matching of Criterion Ratings with Modal Ratings for Each Item by Film and Session (S_1,S_2,S_3). (1 = hit, 0 = miss)*

		Film								
		Meters			Concrete			Ski		
Section	Item	S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
I: Production	1.	0	0[b]	0[b]	0[b]	0	0[b]	1[b]	0	0
	2.	1	1[b]	1[b]	1[b]	1	1[b]	1	1	1
	3.	1	1	1	1	1	1	1	1	1
	4.	1	1	1	1	1	1	1	1	1
II: Pictorial	5.	1	1	1	1	1	1	1	1	1
	6.	1	1	1	0	0	0	1	1	1
	7.	1	1	0	1	1	1	1	1	1
	8.	1	1	1	1	1	1	1	1	1
	9.	1	1	1	1	1	1	1	1	1
	10.	1	1	1	1	1	1	0	0	0
	11.	1	1	1	1	1	1	0	0	0
	12.	1	1	1	1	1	1	1	1	1
	13.	1	1	1	1	1	1	0	0	0
III: Verbal	14.	1	1	1	1	1	1	1	1	1
	15.	1	0	0	1	1	1	1	1	1
	16.	1	1	1	0	0	0	0	1	0
	(17.)[a]	(0)	(0)	(0)	(1)	(1)	(1)	(0)	(1)	(0)
	18.	1	1	1	1	1	1	1	1	1
	19.	0	0	0	1	1	1	1	1	1
	20.	1	1	1	1	1	1	1	1	1
	21.	1	1	1	0	0	0	0	1	1
IV: Organization	(22.)[a]	(1)	(1)	(1)	(1)	(1)	(1)	(0)	(1)[b]	(1)
	23.	1	0	0	0	0	0	1	1	1
	(24.)[a]	(1)[b]	(0)	(0)	(0)	(1)	(1)	(1)[c]	(0)	(1)
	Totals	19	17	16	16	15	16	16	17	16
	Film Totals	Meters 52 (83%)			Concrete 47 (75%)			Ski 49 (78%)		
	Session Totals	Session 1 51 (81%)			Session 2 49 (78%)			Session 3 48 (76%)		
	Grand Total	148 (78%)								

[a] Dropped Item

[b] Bimodal Item

[c] Trimodal Item

*"hit" indicates the rater's modal rating for that item was equal to the criterion rating.

Discussion

Dropped Items

Based on our analysis relevant to the first question, it was decided to delete three questions from the instrument. These were items 17, 22 and 24.

Items 22 (Are the purposes of the presentation introduced clearly in the beginning?) and 24 (Is the presentation well organized?) may have been difficult for the raters because they involved rating film organization on a 5-point scale. Though raters were instructed to "rate stringently" on organizational points because of the need of NTID students to have extremely well organized presentations, there was a high degree of disagreement among raters. Possibly because of the subjectivity of these items, raters had difficulty placing judgments on a 5-point scale.

Item 17 (Number of technical terms or new vocabulary introduced and defined) was difficult because raters not only had to pick up the number of technical terms used (cf. item 16), but they also had to note whether or not the terms were defined in the film. Undefined terms (e.g, "psi" or "pounds per square inch") may not bother a highly educated rater, but would pose difficulty for an NTID student who had less experience with such terms before entering college.

Section and Item Characteristics

The following is a discussion of the results (by section) of each individual item. Characteristics of individual items that may have led to the observed results are also discussed.

Section I: Production Characteristics (4 items). Overall, the 2/3 level of agreement among raters on all ratings for the four items was reached 69% of the time for this section of the checklist. (These data are shown in Table 3 in Appendix H.)

Modal agreement (the percent of ratings that equalled the overall mode for that item) within the dimension was 85%. (See Table 4 in Appendix H.)

Overall agreement of the modal ratings with the criterion ratings was 75% for this dimension (see Table 3, Appendix H).

Item 1 (Approximate age-group of viewers) was the most difficult in this section. Since NTID serves adults, item 1 should be modified to address the question "Does the content of the film appear to be directed to adults (post high school) or is it clearly a child's show (i.e., with all child actors and childhood themes)?" For school district use, perhaps the item should be retained since raters selecting films would have familiarity with curriculum materials appropriate for various grade levels.

For item 2 (Type of presentation) the 2/3 level of agreement was not reached for ratings on one film, and two of three times on another. Raters reported having had difficulty distinguishing item 2 from item 3 (General instructional purpose.) The Rater's Guide should be updated to emphasize that item 2 addresses format and item 3 addresses content of the presentation.

Section II. Pictorial Characteristics (9 items). This section had the best overall interrater agreement of 84% of the ratings at the 2/3 level. The section had an overall modal agreement of 83%.

Interrater agreement was high for all items except for item 13 (How well do the pictures alone convey the message?) on one film (Ski), the modal ratings of which failed to agree with the criterion as well. This item required the rater to judge how adequate the pictorial presentation was on its own (without verbal information). One can speculate that raters varied considerably in their stringency since normally hearing adults probably do not usually parse film viewing into verbal and pictorial components.

Modal agreement on three other items did not agree with the criterion for one film (items 6, 10, and 11). The problem with item 6 (Cartoon- like photographic setting.) was clear. The raters failed to detect the cartoon animation segments that appeared in the film (Concrete) and rated the segments as drawings. This suggests that a variety of cartoon animation examples should be shown to raters during training.

On item 10 (Are close-ups provided for detail when needed?) raters were supposed to judge the use of close-up photography. In this case, the criterion rater was more strict, basing ratings on quantitative data rather than on overall impression (Ski film).

Again, for item 11 (Is color used when needed?), the criterion rater was more strict in judgment; this time for the use of color (Ski film). The raters possibly did not detect the verbal reference to a red compass needle which was not shown in a view closely enough to see its color. This detection failure may account for rater-criterion discrepancies for both items 10 and 11.

Section III. Verbal Characteristics (7 items). Interrater agreement reached the 2/3 level for 68% of the items in this section. Agreement of the item modes with the criterion was 78% overall.

Five of the seven items in this section have one instance for which interrater agreement did not reach 2/3 for a film or did not match the criterion rating on a film. These will be discussed in order.

For item 14 (Concrete), verbal information load, modal agreement with criterion was 100%, but the 2/3 level of interrater agreement was not reached. The raters were split between "high" and "medium" ratings on this item with the mode at the "high" end. Thus, there is little actual difficulty with this item.

Agreement among raters on item 15 (English skills required) did not reach the 2/3 level for one film (Meters). Modal ratings matched the criterion ratings for only one viewing. The problem with this item may be related to that discussed with item 1, the difficulty raters had in judging age-appropriateness of a film. Debriefing comments of raters confirmed this conclusion.

Item 16 (Technical vocabulary assumed of viewer), the detection of defined technical terms, did not reach the 2/3 level of agreement or a match of modal ratings with the criterion for one film (Concrete) and had agreement on one viewing for both

indices on another (Ski). This item may have a problem akin to the one which lead to dropping item 17 (Number of technical terms introduced and defined), that of the ability of a highly educated rater to detect a technical term or new vocabulary. Much care was taken to explain item 16 and its companion in the Rater's Guide and in training, but a script may really be necessary for this task under normal conditions. In contrast, technical terms (units of measurement) in Meters were very salient. They were written apart from the caption and were presented and defined in a fixed, predictable pattern.

Item 19 (Distinctiveness of captions for vocabulary or important concepts), had no match of modal agreement with the criterion ratings and only one viewing with interrater agreement at the 2/3 level for one film (Meters). The reason for this was clear. Raters were told to treat verbal material which was embedded in the picture (e.g., in book titles, road signs, and "super-imposed writing" over the picture) as pictorial information for their ratings. Though we had reasons for this (to judge the adequacy of the captions), the confusion generated by the instructions suggests that we were wrong. After all, the print used for the units of measurement in Meters was highly distinct. Future versions of the checklist should not distinguish between verbal information presented in captions and that presented in the pictures (or before captions were added).

Item 21, (Caption legibility) had modal agreement that did not match the criterion rating and interrater agreement that reached the 2/3 level for only one viewing of one film (Concrete). The reason for this disparity is unclear. Though raters commented about the problems in reading many captions in the film, most raters rated it as "sometimes" rather than "yes". Again, the criterion rater was more strict on this item, which suggests that raters should rate more critically on caption legibility.

Section IV. Organizational Characteristics. Agreement for items in this section is insufficient for discussion of individual items (although item 23, on summaries, survived the cut for inclusion in analysis). The difficulty in rating with the 5-point scales used in

this section has been discussed, as has the subjectivity of judging organization. We believe that what these items purport to assess--organization--is highly important in education of deaf students. Because ratings tended to cluster among adjacent points, we suggest that the use of a 3-point scale will enable these items to be used in future studies.

Suggestions for Revisions and potential users

Based on this study and a review of this paper by several colleagues, a variety of changes were made in the instrument. The revised instrument is attached as Appendix I.

In addition, the following changes were made to the original checklist and are reflected in the Revised Film Evaluation Checklist:

1. Change name of Section I to "Format Characteristics," since "Production" is often connoted to mean "Pictorial Characteristics."
2. Item 1 on age-appropriateness may have to be modified slightly depending on the potential audience of the rated films.
3. Drop item 3 (general instructional purpose) since purposes of a presentation may vary greatly in different educational contexts.
4. Modify Note on item 13 to read "...presented by voice or captions and include words embedded in pictures..."
5. Drop item 17 because it had poor reliability characteristics.
6. Include Section IV. (Organization) items, but on a three-point scale.

Limitations and Suggestions for Future Research

Any instrument of this type is only valuable for the user if it provides the needed information. Based on our work, we feel this instrument provides a useful tool for describing and categorizing certain educationally relevant characteristics of films. However, any potential user must examine the items to see if the information provided is useful.

Our nine raters were able to substantially agree on a film's characteristics after only one viewing. The study might profitably be repeated, however, with a greater number of raters.

It is important to note that instructional films are rarely, if ever, presented in isolation. The context in which the film is presented (i.e., how it is introduced by the instructor or fits into a course plan, the type of student watching the film, whether or not the students will be tested on the content) all may have as much or more impact on student learning than the characteristics of the film itself. Anyone intending to use the Film Evaluation Checklist in a research project should, of course, collect information on such context variables.

Further research on the "organizational characteristics" section of the checklist is needed, particularly in view of the noted difficulty with 2 of the 3 items in that section.

Finally, the particular characteristics that we designated as "instructionally important" were drawn from the literature and from discussions with many people. Thus, while the instrument has face validity, relationships with outcome criteria such as student learning and satisfaction has not been examined. This should be done in the future.

Summary

This study was directed toward the following questions about the reliability characteristics of the Film Evaluation Checklist:

1. Would trained raters rate the items on this instrument with a reasonable degree of consensus over three viewings of a film?
2. Would agreement among the nine raters increase over three viewings of a film?
3. Would the ratings approximate criterion ratings made by a skilled rater?

The first question, of consensus, was addressed by dropping items for which six raters did not agree for at least three ratings. Three items were dropped from analysis,

although recommendations were made to modify the response format for the two items on organization. After these items were eliminated, it was found that six or more raters agreed for 74% of their ratings, a satisfactory level of agreement.

To answer the second question, of agreement over time, agreement within and among raters was computed by comparing each item rating to its corresponding mode. Though it was expected that agreement would increase over viewings, ratings were consistent. Overall agreement among raters was slightly higher for Session 1, suggesting that films may be reliably evaluated after a single viewing.

The third question, of convergence with criterion ratings, was addressed by comparing modal ratings for each item to criterion ratings made by a skilled rater. This comparison yielded 78% agreement between the trained raters and the criterion rater, suggesting that they were rating on the same basis. This accuracy was also stable across viewings (though slightly higher for Session 1). These results strengthen the conclusion to the second question, stating that not only do raters agree after a single viewing, they agree with the criterion rater who trained them.

In summary, a film evaluation instrument of acceptable reliability and accuracy has been developed. It was shown that first viewings after a 2-hour training session yielded ratings as reliable and accurate as did subsequent rating sessions.

For implementation as an evaluation instrument, the Film Evaluation Checklist may be used for both captioned and uncaptioned films, since most Verbal Characteristics items address the amount and level of verbal content. This study did not evaluate, however, raters' ability to rate aural and captioned verbal information simultaneously. The Film Evaluation Checklist, with the modifications discussed earlier, should serve as a balanced instrument for evaluators who need to routinely select films with specific characteristics or for researchers who need to categorize films by certain characteristics for use in research studies. It should be noted that the use of the checklist would not necessarily be limited to deaf education.

Reference Notes

1. Marantz, S., & Dowaliby, F. Film versus lecture mode of instruction as related to imageability. Paper presented at the meeting of the American Educational Research Association, New Orleans, February, 1973.

2. Braverman, B. B., Harrison, M. F., Dowaliby, F. J., & Bowker, D. O. The effect of reading level, and visual display, on learning from captioned instruction. Paper presented at the meeting of the American Educational Research Association, Toronto, 1978.

3. Braverman, D. The effects of caption rate and word density on comprehension of a video presentation. Unpublished paper. National Technical Institute for the Deaf, Rochester, NY, 1981.

4. Dowaliby, F. J., Burke, N. E., & McKee, B. G. Validity and reliability of a learning style inventory for postsecondary deaf individuals. Paper presented at the annual meeting of the American Educational Research Association, Boston, April, 1980.

References

Braverman, B.B., & Hertzog, M. The effects of caption rate and language level on comprehension of a captioned video presentation. American Annals of the Deaf, 1980, 125, 943-947.

Cronbach, L. J. & Snow, R. E. Aptitudes and Instructional Methods. New York: Irvington Publishers, Inc., 1977.

Cross Country Ski Technique. Douglas Sinclair Film Productions, 1973. (Film)

Elements, Compounds and Mixtures. Educational Film Services, 1961 (Film).

Hochberg, J., & Brooks, V. The perception of motion pictures. In E. C. Carterette & M. P. Friedman (Eds.), Handbook of perception (Vol. X.), New York: Academic Press, 1978.

Johnson, D. D. Communication characteristics of a young deaf adult population: techniques for evaluating their communication skills. American Annals of the Deaf, 1976, 121, 409-424.

Koran, M., Snow, R. E., and McDonald, F. Teacher aptitude and observational learning of a teaching skill. Journal of Educational Psychology, 1971, 62, 219-228.

McKee, B. G. & Hausknecht, M. A. Classroom assessment techniques for hearing impaired students: A literature review. Dimensions, 1980, 1, 9-15. (ERIC Document Reproduction Service No. ED 161 937).

Meters, liters, and kilograms. Perennial Education, 1974. (Film)

Parasnis, I., and Long, G. L. Relationships among spatial skills, communication skills, and field independence in deaf students. Perceptual and Motor Skills, 1979, 49, 879-887.

Principles of quality concrete. Portland Cement Association, 1969. (Film)

Quigley, S., Wilbur, R., Power, D., Montanelli, D., & Steinkamp, M. Syntactic structures in the language of deaf children. Institute for Child Behavior and Development, University of Illinois at Urbana-Champaign, 1976.

Salomon, G., & Snow, R. E. The specification of film attributes for psychological and educational research purposes. Audio-Visual Communications Review, 1968, 16, 225-244.

The time of your life. Cally Curtis Co., 1974. (Film)

Wilbur, R. B., Explanation of Deaf Children's Difficulty with Certain Syntactic Structures of English. The Volta Review, 1977, 79 (2), 85-92.

Witkin, H., Moore, C., Goodenough, D., & Cox, P. Field-dependent and field-independent cognitive styles and their educational implications. Review of Educational Research, 1977, 47, 1-64.

Appendix A

History of the Visual Media Research Project

History of the Visual Media Research Project (1978-80)

The Film Evaluation Checklist was developed as part of a research project entitled the "Visual Media Research Project." The project was initiated in 1978 by what was then the Division of Curriculum Development and Evaluation. The Visual Media Research Project was conceived as a division wide five-year aptitude-treatment interaction (ATI) study designed to find interactions among student characteristics (e.g. achievement measures, aptitude measures, demographic variables, and cognitive styles), film characteristics, instructional variables, and film-based learning. Once evaluation instruments and procedures were developed, all instructional films used in the Institute and student learning from these films were to be evaluated over a two- or three year period.

The major goal of the Visual Media Research Project was to delineate instructional film characteristics that would aid in the selection, rental, purchase, captioning, or production of films at NTID that would be most suitable for deaf learners. Since much Institute effort is given to these functions, it was reasoned that the isolation of film characteristics which maximize learning by NTID students would be a worthwhile contribution to deaf education. It was also our intention that the project would serve a short term materials evaluation need by providing immediate feedback concerning learning and attitudinal outcomes of a given piece of instructional media.

Two major evaluation instruments, the Film Evaluation Checklist and the Learning Style Inventory (Dowaliby, Burke & McKee, Note 4) were developed during the two active years of the project.

The Project Team at the inception of the Visual Media Research Project was:

Project Co-Leaders:

Fred J. Dowaliby & Barbara G. McKee, Research Associates in Curriculum Development & Research (now in Educational Research & Development)

Project Assistant

Nancy E. Burke, University of Rochester doctoral student on a research assistantship with F. J. Dowaliby and B. G. McKee.

Management Consultants:

Frank Argento, Director of the Division of Curriculum Development & Evaluation

Barry Cronin, Chair of Media Instructional Television

Roberta Stevens, Chair of Media Services

Harold Farneth, Chair of Curriculum Development & Research

In the Winter of 1980, the Project lost the support from the Management Consultants through divisional reorganization, personnel changes, and resulting policy changes. In a memo dated February 15, 1980, Jim Carroll, the Director of Media Development & Services, officially terminated the project. It was decided to complete development work on the Film Evaluation Checklist and the Learning Style Inventory in the Department of Educational Research & Development.

Appendix B

Project Consultants and Reviewers

List of People Consulted on Project

Person

Linda Bardenstein, Curriculum Materials Specialist

Jo Bausch, Curriculum Materials Specialist

Roland Blake, Instructional Developer

Barbara Braverman, Research Associate

Barry Cronin, Chairperson, NTID, Instructional Television Department

James DeCaro, Instructional Developer

Judy Egelston-Dodd, Instructional Developer

Steven Marantz, Marantz Media

Michael Smith, Instructional Developer

Richard E. Snow, Stanford University

Roberta Stevens, Chairperson, NTID, Media Service Department

Reviewers

Barbara Braverman, Research Associate

Wayne Garrison, Research Associate

John Lutz, Research Associate

Appendix C

Film Evaluation Checklist

VISUAL MEDIA RESEARCH PROJECT
FILM EVALUATION CHECKLIST
JULY 1980

FILM EVALUATION CHECKLIST
DEVELOPED BY

NANCY E. BURKE
FRED J. DORALIBY
BARBARA B. MCKEE

NTID DEPARTMENT OF EDUCATIONAL RESEARCH AND DEVELOPMENT
IN COOPERATION WITH NDB

VMRP
VMRP
VMRP
VMRP
VMRP
VMRP
VMRP
VMRP
VMRP
VMRP

Rater _____

Viewing: 1st _____

2nd _____

3rd _____

July 1980 Version

Rater Study

VMAP

FILM EVALUATION CHECKLIST

I. PRODUCTION CHARACTERISTICS

Title _____

Running Time _____. Copyright Date _____.

Format (check one):

a. Film _____

b. Videotape _____

c. Filmstrip/slides _____

Photography (check one):

a. Color _____

b. Black and White _____

Is the presentation captioned? (check one): Yes _____ No _____

Is manual communication used by people in the presentation? (check one): Yes _____ No _____

1. Appropriate age-group of viewers. (Please check the lowest suitable age-group for viewing the presentation):

a. College + (adult) _____

b. High school (gr. 10-12) _____

c. Middle school (gr. 7-9) _____

d. Grade school _____

2. Type of presentation (check the best description):

a. Lecture _____

b. Demonstration ("how to") _____

c. Documentary (depictions of real settings, events, interviews of participants). _____

d. Discussion (panel or roundtable) _____

e. Dramatization (story). _____

3. General Instructional Purpose (check the best description):

a. Information ("cognitive") _____

b. Skill demonstration ("motor skill," "how to") _____

c. Motivation ("affective") _____

d. Entertainment _____

4. Is the presentation current? (check one):

a. Seems current and/or historically appropriate. _____

b. Seems a bit out of date, but does not detract from educational content. _____

c. Out of date to the point that it would detract from educational content. _____

II. PICTORIAL CHARACTERISTICS

Photographic Setting	Check which one applies for each number: Primary	Secondary	Absent
5. Life-like settings	_____	_____	_____
6. Cartoon-like	_____	_____	_____
7. Charts, graphs, non-cartoon like drawings	_____	_____	_____

Use of Motion	Check which one applies for each number: Primary	Secondary	Absent
8. Motion/Animation Photography (not zooms or pans of stills)	_____	_____	_____
9. Time lapse/slow motion Photography	_____	_____	_____

10. Are close-ups provided for detail when needed? (Check one):

Yes _____ Sometimes _____ No _____ N.A. _____

11. Is color used when it is needed in getting the message across? (Check one):

Yes _____ Sometimes _____ No _____ N.A. _____

12. Are pictorial examples of objects/concepts which are discussed in verbal form given when needed?

Yes _____ Sometimes _____ No _____ N.A. _____

13. How well do the pictures alone convey the message? (check one):

a. Verbal information unnecessary _____

b. Message is often comprehensible. _____

c. Would have to guess often. _____

d. Verbal information essential. _____

NOTE: Verbal Information is that which is presented by voice or captions. (Does not include words embedded in pictures or "written" c

III. VERBAL CHARACTERISTICS

14. Verbal information load. (check one):

High	Medium	Low
_____	_____	_____

15. English skills required (check one):

High	Medium	Low
_____	_____	_____
(College)	(High School)	(Primary)

16. Technical vocabulary <u>assumed</u> of viewer (check one):

Adv.	Inter.	Beg.
_____	_____	_____

NOTE: Terms used that are not defined.

17. Number of technical terms (or new vocabulary) introduced and defined (check one):

Many	Moder.	Few
_____	_____	_____

18. <u>If Captioned</u>: Does screen positioning of captions show which actor is speaking? (If film has only one speaker, check "Yes.") Check one:

Yes	Sometimes	No
_____	_____	_____

19. <u>If Captioned</u>: Are captions of new concepts and vocabulary distinguishable from <u>rest</u> of text? (check one):

Distinct	Marginally Distinct	<u>Not</u> Distinct
_____	_____	_____
(all caps; underlined or out of text)	(Cap 1st letter or has slightly different font)	(Close to or same as text)

20. <u>If Captioned</u>: Consistency of ways used to caption new concepts/vocabulary (check one):

Consistent	Moderately Consistent	Inconsistent
_____	_____	____

21. <u>If Captioned</u>: Does pacing of shots or captions interfere with <u>reading</u> of captions? (Applies to rate of cutting, action within shots, or busy captions.) Check one:

Yes _____ Sometimes _____ No _____

IV. ORGANIZATIONAL CHARACTERISTICS

	Very Clearly	Clearly	Sort of Clearly	Unclearly	Very Unclearly
22. Are the purposes of the presentation introduced clearly in the beginning?	1	2	3	4	5
23. Are summaries given when appropriate in the presentation?	1	2	3	4	5
24. Is the presentation well organised?	1	2	3	4	5

Additional comments:

Appendix D

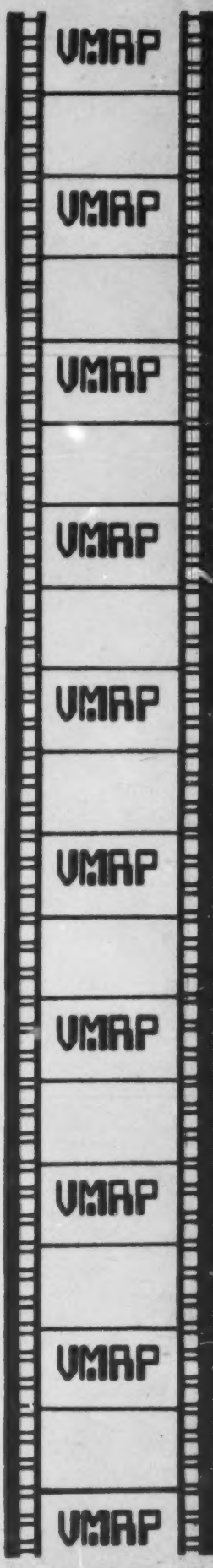
Master Criterion Rating Checklist

VISUAL MEDIA RESEARCH PROJECT
FILM EVALUATION CHECKLIST
JULY 1980

FILM EVALUATION CHECKLIST
DEVELOPED BY

NANCY E. BURKE
FRED J. DONALIBY
BARBARA B. MCKEE

NTID DEPARTMENT OF EDUCATIONAL RESEARCH AND DEVELOPMENT
IN COOPERATION WITH NDS



Rater _____

Viewing: 1st _____

2nd _____

3rd _____

July 1980 Version

Rater Study

VMAP

FILM EVALUATION CHECKLIST

I. PRODUCTION CHARACTERISTICS

Title _____

Running Time _____. Copyright Date _____.

Format (check one):

a. Film _____

b. Videotape _____

c. Filmstrip/slides _____

Photography (check one):

a. Color _____

b. Black and White _____

Is the presentation captioned? (check one): Yes _____ No _____

Is manual communication used by people in the presentation? (check one): Yes _____ No _____

1. Appropriate age-group of viewers. (Please check the lowest suitable age-group for viewing the presentation):

a. College + (adult) _____

b. High school (gr. 10-12) _____

c. Middle school (gr. 7-9) _____

d. Grade school _____

2. Type of presentation (check the best description):

a. Lecture _____

b. Demonstration ("how to") _____

c. Documentary (depictions of real settings, events, interviews of participants). _____

d. Discussion (panel or roundtable) _____

e. Dramatization (story). _____

3. General Instructional Purpose (check the best description):

a. Information ("cognitive") _____

b. Skill demonstration ("motor skill," "how to") _____

c. Motivation ("affective") _____

d. Entertainment _____

4. Is the presentation current? (check one):

a. Seems current and/or historically appropriate. _____

b. Seems a bit out of date, but does not detract from educational content. _____

c. Out of date to the point that it would detract from educational content. _____

II. PICTORIAL CHARACTERISTICS

Photographic Setting	Check which one applies for each number: Primary	Secondary	Absent
5. Life-like settings	_____	_____	_____
6. Cartoon-like	_____	_____	_____
7. Charts, graphs, non-cartoon like drawings	_____	_____	_____

Use of Motion	Check which one applies for each number: Primary	Secondary	Absent
8. Motion/Animation Photography (not zooms or pans of stills)	_____	_____	_____
9. Time lapse/slow motion Photography	_____	_____	_____

10. Are close-ups provided for detail when needed? (Check one):

Yes _____ (2 needed) Sometimes _____ (2-3) No _____ (4+ needed) N.A. _____

11. Is color used when it is needed in getting the message across? (Check one):

Yes _____ (always) Sometimes _____ (yes with 1 exception) No _____ (2+ needed instances) N.A. _____ (B&W film where color would not add - or color not important in color film beyond aesthetic value)

Note: "Yes" and "N.A." categories are scored together.

12. Are pictorial examples of objects/concepts which are discussed in verbal form given when needed?

Yes _____ (80%) Sometimes _____ (20%+) No _____ (20%) N.A. _____

13. How well do the pictures alone convey the message? (check one):

a. Verbal information unnecessary _____

b. Message is often comprehensible. _____

c. Would have to guess often. _____

d. Verbal information essential. _____

NOTE: Verbal Information is that which is presented by voice or captions. (Does not include words embedded in pictures or "written" on.)

III. VERBAL CHARACTERISTICS

14. Verbal information load. (check one):

High	Medium	Low
15+	6-14	0-6

Note: (captions/min.)

15. English skills required (check one):

High	Medium	Low
(College) 12th grade	(High School) 7-12	(Primary) 7

Note: (Dale Chall x grade level)

16. Technical vocabulary <u>assumed</u> of viewer (check one):

Adv.	Inter.	Beg.
20%	10% to 20%	10%

(Percentage of unique technical terms not defined, as defined below)

NOTE: Terms used that are not defined.

17. Number of technical terms (or new vocabulary) introduced and defined (check one):

Many	Moder.	Few
2/min	1 to 2/min	1/min

(Index should be a "term/minute" measure. Definition can be verbal, visual, or both.)

18. <u>If Captioned</u>: Does screen positioning of captions show which actor is speaking? (If film has only one speaker, check "Yes.") Check one:

Yes	Sometimes	No

19. <u>If Captioned</u>: Are captions of new concepts and vocabulary distinguishable from <u>rest</u> of text? (check one):

Distinct	Marginally Distinct	<u>Not</u> Distinct
(all caps; underlined or out of text)	(Cap 1st letter or has slightly different font)	(Close to or same as text)

20. <u>If Captioned</u>: Consistency of ways used to caption new concepts/vocabulary (check one):

Consistent	Moderately Consistent	Inconsistent

21. <u>If Captioned</u>: Does pacing of shots or captions interfere with <u>reading</u> of captions? (Applies to rate of cutting, action within shots, or busy captions.) Check one:

Yes	Sometimes	No
5%	1-5%	1%4-

USE OUTLINES HERE

IV. ORGANIZATIONAL CHARACTERISTICS

	Very Clearly	Clearly	Sort of Clearly	Unclearly	Very Unclearly
22. Are the purposes of the presentation introduced clearly in the beginning?	1	2	3	4	5
23. Are summaries given when appropriate in the presentation?	1	2	3	4	5
24. Is the presentation well organized?	1	2	3	4	5

Additional comments:

Appendix E

Description of Films and Criterion Ratings Assigned to Films

Cross-Country Ski Technique (1973)

Demonstrates skills useful in cross-country skiing. Visual examples are given of all verbally stated techniques. (Some in slow motion.) Summaries are given. Verbal demands moderate.

Criterion Ratings:

Item

1. Grades 7 - 9
2. Demonstration
3. Skill Demonstration
4. Current presentation
5. Life-like setting: primary
6. Cartoons: absent
7. Charts: absent
8. Motion photography: primary
9. Time-lapse/Slow motion photography: secondary
10. Close-ups: sometimes
11. Color: sometimes
12. Pictorial Examples: yes
13. Pictures alone? must guess
14. Verbal load: medium
15. English skills: medium
16. Assumed terms: intermediate
17. New terms: few
18. Caption possible for speaker: yes
19. Captions of concepts: same as text
20. Consistency of captions for concepts: consistent
21. Caption interference: no
22. Introduction (3) sort of clearly
23. Summaries (2) clearly
24. Organized? (3) sort of clearly

Meters, Liters, and Kilograms (1974). Metric education film intended for late elementary school - early middle school audiences. (Older children act in film.) Metric terms introduced one by one with visual examples of concrete objects of corresponding sizes and measurement procedures. Moderate verbal demands. Organizational problem in film is that no transitions between linear, square (area), and cubic (volume, density) concepts are provided. (Concepts actually too difficult for intended audience.)

Criterion Ratings:

Items:

1. Grade school
2. Demonstration
3. Information
4. Current presentation
5. Life-like setting: primary
6. Cartoon: absent
7. Charts: secondary
8. Motion photography: primary
9. Time lapse/slow motion photography: secondary
10. Close-ups: yes
11. Color: N.A.
12. Pictorial examples: yes
13. Pictures alone: often comprehensible
14. Verbal load: medium
15. English skills: medium
16. Assumed terms: beginner
17. New terms: few
18. Caption positions for speaker: yes
19. Captions of concepts: same as text
20. Consistency of captions for concepts: consistent
21. Caption interference: no
22. Introduction (2) clearly
23. Summaries (3) sort of clearly
24. Organized? (3) sort of clearly

Principles of Quality Concrete (1969). Introduction to materials and procedures used in preparing concrete. Technical film. High rates of verbal information in text and on charts and graphs. Good visual examples.

Criterion Ratings

Items

1. Grade: College +
2. Demonstration
3. Information
4. Current presentation
5. Life-like setting: primary
6. Cartoons: secondary
7. Charts: secondary
8. Motion photography: primary
9. Time-lapse/slow motion: absent
10. Close-ups: yes
11. Color: yes
12. Pictorial examples: yes
13. Pictures alone? verbal essential
14. Verbal load: high
15. English skills: medium
16. Assumed terms: advanced
17. New terms: moderate
18. Caption positions for speakers: yes
19. Captions of concepts: same as text
20. Consistency of captions for concepts: consistent
21. Caption interference: yes
22. Introduction (2) clearly
23. Summaries (4) unclearly
24. Organized? (3) sort of clearly

Appendix F

Rater's Guide to the Film Evaluation Checklist

VISUAL MEDIA RESEARCH PROJECT

Rater's Guide to the Film Evaluation Checklist

July 1980

Nancy E. Burke, Fred J. Dowaliby & Barbara G. McKee
NTID Department of Educational Research & Development

During the past two years, a checklist has been under development for the purpose of evaluating educational films at NTID. This instrument was developed for several reasons.

The first is the increased interest in captioning films for the deaf. As more captioned films become available, methods of comparing films must be developed.

Also, the checklist may be used to evaluate the feasibility of captioning a film. Verbal, pictorial, and organizational characteristics of films may be rated with this checklist. For example, a highly verbal film may be rejected in favor of a highly pictorial film on the same subject. It should be noted that no instruments of this type exist in the educational field for use with uncaptioned films. The film evaluation checklist may be used for this purpose as well.

Thirdly, we recognize the special needs and learning styles of deaf individuals which are relevant to picking up information from films. Certain verbal, pictorial, and organizational characteristics of instructional films may interact with certain student learning characteristics. The film evaluation checklist may serve as a research tool to isolate film characteristics which interact with student learning styles.

The purpose of the present study is to evaluate agreement among trained raters after one, two, and three ratings of three different instructional films. Three films will be rated once per week for three weeks. We are also interested in finding out how ratings agree with master ratings made by one of the checklist authors. Each film was studied in depth by that person.

We offer several ground rules for raters participating in this study:

(1) Each rater must see each of the three films once per week at a single sitting and in a specified order. Each film will be viewed for a total of three times during the study.

(2) Rating forms are turned in after each rating session to Nan Burke in ER&D.

(3) Raters may not discuss any of the films with other raters during the study. We want ratings to be as independent as possible.

(4) Please appear at scheduled rating sessions. Nan Burke is contracted for only two days per weeks.

I. PRODUCTION CHARACTERISTICS.

This section is for rating the formats and general purposes of the presentation.

(2-3) Most films are a mixture of these characteristics. We want raters to select the best description of those offered.

II. PICTORIAL CHARACTERISTICS.

This section is for evaluating the photography and the adequacy of the pictorial content.

(5-7) Photographic Setting. Ratings are made for the kinds of artwork used for the presentation.

(8-9) Use of Motion. All films use motion picture photography; filmstrips do not. Films vary in their use of motion picture photography, however. (8) Films are sometimes made using pans and zooms on still photographs, i.e., on low budget commercials or historical films. Here, the camera lens is moving, but not the photograph or drawing.

(9) The speed used in filming is sometimes varied to make information for an event more salient for an audience. Timelapse photography is used to speed up an event. For example, a flower that takes 20 hours to bloom may be shown to bloom in 20 seconds. Slow motion photography, however, slows down an event. Televised sports coverage often shows replays in slow motion to illustrate the intricacies of important plays.

(10) Close-up photography is often used to provide important details for viewers. For example, in one of the training films, properties of various substances are shown in close-up. Close-ups of people's faces in dramatized scenes often provide information about emotions, attitudes, etc.

(11) Color is sometimes important in instructional films, e.g., in use of color for graphic composition, chemical reactions, and for learning to wire circuits. Some kinds of viewers require color more than others for such films.

(12) One of the advantages of a motion picture over a textbook is that visual examples are given of important concepts. For an audience with reading problems, the importance of the visual examples in a captioned film cannot be overstated.

(13) Outside of the "silent film" genre, few films are adequate without sound. How understandable is the film without sound before captioning?

III. VERBAL CHARACTERISTICS

Ratings in this section are indices of the amount and difficulty of verbal material presented in the film.

(14) This item is meant to secure an overall impression of the sheer amounts of verbal information presented. Is there time to watch both the pictures and the captions?

(15) Addresses the grade level of the text presented.

(16-17) Technical terms are terms with specific meanings in certain fields--e.g., reaction or centigrade. Some terms are presented and defined. Others are implicit and are not formally defined.

Please ignore technical terms incidental to the film, e.g., where in a management film, an employee gives her boss a "snow job" in "computerese".

(18) Captioning positions are often used to indicate right, left, center, and off-camera voices when several voices are used.

(19) Sometimes important words are underlined, capitalized, are in a different type, or are set out of text. Is this done?

(20) Is #19 done consistently?

(21) Sometimes there are problems with captions which make them hard to read. A caption may appear over an animated chart or over somebody's name. Changes in the picture (cuts, motion) may also make captions difficult to read.

IV. ORGANIZATIONAL CHARACTERISTICS.

Research here at NTID has shown that deaf students require much explicit structure in the organization of information presented to them. This section is to evaluate the clarity and organization of information presented in films.

(22) Do you know what this film is going to be about after watching for the first 2 or 3 minutes? How much of the organization of the film is tipped off at the beginning?

(23) Sometimes you feel the film should make a short summary at certain points in the film. Does it? Or do you feel compelled to rewind?

(24) Are you able to make a simple outline of the presentation when it is complete? Did you have a sense of "what was going to come next" throughout the film, i.e., a sense of logical progression?

Appendix G

Outline for Film Rater Training

VISUAL MEDIA RESEARCH PROJECT

Film Evaluation Validation Project

*OUTLINE FOR FILM RATER TRAINING (For July 1980)

Ten NTID faculty members have agreed to rate three films at one-week intervals. Each film will be rated three times: once during each weekly 2-hour rating session. Raters will be trained during a preliminary 2-hour session to take place the second week of July.

I. Purpose of Project:

A. Increasing availability of captioned films for instruction of deaf individuals.

B. Now have a better understanding of special characteristics of learning styles and needs of deaf individuals.

C. Felt there was a need for an easily used film evaluation instrument to be used to rate captioned and uncaptioned films. Institutions serving the needs of deaf students need a way to sift through and evaluate films available to them. For example, there are a number of birth control films available. How do you choose the best film available?

D. Have developed a checklist that may serve this purpose. One of us has made an in-depth evaluation of the three films you will rate. We wish to find out how many viewings (up to 3) it will take for your ratings to converge with ours.

II. Role of Raters.

A. Two-hour training session (present session) to learn to rate films.

B. Rating session I. (Third week of July) Rate each of three films after first viewing. (2-hour session)

C. Rating session II. (Fourth week of July)

D. Rating session III. (Fifth week of July)

III. Overview of Checklist Organization.

A. Production Characteristics--e.g., When the film was produced; whether it is in color or black & white, etc.

B. Verbal Information--e.g., vocabulary levels required, what caption characteristics exist, etc.

C. Pictorial Information--e.g., use of color, close-ups, and visual examples when appropriate.

D. Organizational Characteristics--e.g., Are there introductions, summaries, etc. when appropriate?

IV. How to Rate Films Using the Checklist.

A. Production Characteristics: define all points, give examples.

B. Verbal Information: define all points. Show "film clip" examples when appropriate.

C. Pictorial Information: define all points. Show "film clip" examples.

D. Organizational Characteristics: define all points. Show examples of intro and summary sequences. Illustrate verbal and pictorial aspects of these.

V. Rating a Sample Film Using the Checklist.

A. Answer questions arising from IV.

B. View sample film (e.g., "Managing your boss")

C. Rate.

D. Discuss rating outcomes.

VI. Preparation for rating sessions.

A. Answer questions

B. Confirm and circulate viewing schedules showing dates, times, location, and ground rules (e.g., that the ratings be <u>independent</u> within and among raters)

Appendix H

Additional Data

Table 1

Agreement of Ratings with Modal Ratings by Rater and Session (S_1, S_2, S_3)

Rater	Session S_1	Session S_2	Session S_3	Rater Totals Across Sessions
1	51 (81%)*	49 (78%)	35 (56%)	135 (71%)
2	56 (89%)	53 (84%)	52 (83%)	161 (85%)
3	50 (79%)	53 (84%)	57 (91%)	160 (85%)
4	47 (75%)	46 (73%)	52 (83%)	145 (77%)
5	55 (87%)	54 (86%)	53 (84%)	162 (86%)
6	49 (78%)	47 (75%)	56 (89%)	152 (80%)
7	53 (84%)	49 (78%)	50 (79%)	152 (80%)
8	46 (73%)	46 (73%)	42 (67%)	134 (71%)
9	50 (79%)	48 (76%)	52 (83%)	150 (79%)
Totals:	457 (81%)	445 (78%)	449 (79%)	1351 (79%)

*Each use of the checklist required 21 ratings by the rater. Since three films were rated each session, the maximum number of times a given rater could agree with the modal rating at each session was 63.

Table 2

Agreement of Ratings with Modal Ratings by Rater and Film

Rater	Film			Rater Totals Across Films
	Meters	Concrete	Ski	
1	42 (67%)*	45 (71%)	48 (76%)	135 (71%)
2	56 (89%)	52 (83%)	53 (84%)	161 (85%)
3	53 (84%)	50 (79%)	57 (91%)	160 (85%)
4	48 (76%)	49 (78%)	48 (76%)	145 (77%)
5	52 (83%)	57 (91%)	53 (84%)	162 (86%)
6	47 (71%)	46 (73%)	59 (94%)	152 (80%)
7	45 (62%)	52 (83%)	55 (87%)	152 (80%)
8	39 (62%)	45 (71%)	50 (79%)	134 (71%)
9	53 (84%)	48 (76%)	49 (78%)	150 (79%)
Totals:	435 (77%)	494 (78%)	472 (83%)	1351 (79%)

*Each use of the checklist required 21 ratings by the rater. Since each film was rated three times, the maximum number of times a given rater could agree with the modal rating for a given film was 63.

Table 3
Number of Items with 2/3 Level of Rater. Agreement (6 of 9 raters) or More by Section for Each Film and Session

Section	Film Totals			Session Totals		
	Meters	Concrete	Ski	1	2	3
Section I: Production[c] (Items 1-4) Grand Total: 25 (69%)	7 (58%)	9 (75%)	9 (75%)	9 (75%)	8 (67%)	8 (67%)
Section II: Pictorial[b] (Items 5-13) Grand Total: 68 (84%)	21 (78%)	25 (93%)	22 (82%)	24 (89%)	22 (82%)	22 (82%)
Section III: Verbal[c] (Items 14-21, except 17) Grand Total: 43 (68%)	14 (67%)	13 (62%)	16 (76%)	14 (67%)	13 (62%)	16 (76%)

[a]Section I has four items, therefore, maximum possible agreement for both film totals and session totals is 12.

[b]Section II has nine items, therefore, maximum possible agreement for both film totals and session totals is 27.

[c]Section III has seven items, therefore, maximum possible agreement for both film totals and session totals is 21.

63

64

Table 4

Number of Individual Ratings Matching Modal Ratings for Each Section by Film and Session

Section	Film Totals			Session Totals		
	Meters	Concrete	Ski	1	2	3
Section I: Production[c] (Items 1-4) Grand Total: 275 (85%)	8) (82%)	90 (83%)	96 (89%)	95 (88%)	86 (80%)	94 (87%)
Section II: Pictorial[b] (Items 5-13) Grand Total: 606 (83%)	19) (80%)	203 (84%)	208 (86%)	204 (84%)	203 (84%)	199 82%)
Section III: Verbal[c] (Items 14-21, except 17) Grand Total: 443 (78%)	13) (73%)	137 (72%)	168 (89%)	143 (76%)	142 (75%)	139 (74%)

[a]Section I has four items, therefore, maximum possible agreement for both film totals and session totals for all nine raters is 108.

[b]Section II has nine items, therefore, maximum possible agreement for both film totals and session totals for all nine raters is 243.

[c]Section III has seven items, therefore, maximum possible agreement for both film totals and session totals for all nine raters is 189.

Table 5

Matching of Criterion Ratings with Modal Ratings for Each Section by Film and Session

Section	Film Totals			Session Totals		
	Meters	Concrete	Ski	1	2	3
Section I: Production[c] (Items 1-4) Grand Total: 27 (75%)	9 (75%)	8 (67%)	10 (83%)	10 (83%)	8 (67%)	9 (75%)
Section II: Pictorial[b] (Items 5-13) Grand Total: 68 (84%)	26 (96%)	24 (92%)	18 (67%)	23 (85%)	23 (85%)	22 (82%)
Section III: Verbal[c] (Items 14-21, except 17) Grand Total: 49 (73%)	16 (76%)	15 (71%)	18 (86%)	16 (76%)	17 (81%)	16 (76%)

[a]Section I has four items, therefore, maximum possible agreement for both film totals and session totals is 12.

[b]Section II has nine items, therefore, maximum possible agreement for both film totals and session totals is 27.

[c]Section III has seven items, therefore, maximum possible agreement for both film totals and session totals is 21.

Appendix I

Revised Film Evaluation Checklist

REVISED
VISUAL MEDIA RESEARCH PROJECT
FILM EVALUATION CHECKLIST
JANUARY 1981

FILM EVALUATION CHECKLIST
DEVELOPED BY

NANCY E. BURKE
FRED J. DONALIBY
BARBARA B. MCKEE

OTIS DEPARTMENT OF EDUCATIONAL RESEARCH AND DEVELOPMENT
IN COOPERATION WITH MDS

VMRP
VMRP
VMRP
VMRP
VMRP
VMRP
VMRP
VMRP
VMRP
VMRP

Rater _____ Revised

Viewing: 1st _____ January 1981

2nd _____

3rd _____

VMRP

FILM EVALUATION CHECKLIST

I. FORMAT CHARACTERISTICS

Title _____ _____.

Running Time _____. Copyright Date _____.

Format (check one):

a. Film _____

b. Videotape _____

c. Filmstrip/slides _____

Photography (check one):

a. Color _____

b. Black & White _____

Is the presentation captioned? (check one): Yes _____ No _____

Is manual communication used by people in the presentation? (check one): Yes _____ No _____

1. Does the content appear to be directed to adults or is it clearly a child's show (i.e., with all child detors and childhood themes)? (Please check the lowest suitable age-group for viewing the presentation):

a. College (adult) _____

b. High school (gr. 10-12) _____

c. Middle school (gr. 7-9) _____

d. Grade school _____

2. Type of presentation (check the best description):

a. Lecture ____

b. Demonstration ("how to"). ____

c. Documentary (depictions of real settings, events; interviews of participants). ____

d. Discussion (panel or roundtable). ____

e. Dramatization (story). ____

3. Is the presentation current? (check one):

a. Seems current and/or historically appropriate. ____

b. Seems a bit out of date but does not detract from educational content. ____

c. Out of date to the point that it would detract from educational content. ____

4. Does the material present minorities, women and handicapped individuals in a positive, non stereotyped manner? (check one): ____

a. Yes, presentation is acceptable. ____

b. Not applicable, above groups do not appear in presentation. ____

c. No, presentation of one or more of the above groups is negative or stereotypical. ____

II. PICTORIAL CHARACTERISTICS

Photographic Setting	Check which one applies for each number: Primary	Secondary	Absent
5. Life-like settings	____	____	____
6. Cartoon-like	____	____	____
7. Charts, graphs, non cartoon-like drawings	____	____	____

Use of Motion	Check which one applies for each number: Primary	Secondary	Absent
8. Motion/Animation Photography (not zooms or pans of stills)	____	____	____
9. Time lapse/ Slow motion Photography	____	____	____

10. Are close-ups provided for detail when needed? (Check one):

Yes _____ Sometimes _____ No _____ N.A. _____

11. Is color used when it is needed in getting the message across? (Check one):

Yes _____ Sometimes _____ No _____ N.A. _____

12. Are pictorial examples of objects/concepts which are discussed in verbal form given when needed?

Yes _____ Sometimes _____ No _____ N.A. _____

13. How well do the pictures alone convey the message? (Check one):

a. Verbal information unnecessary. _____

b. Message is often comprehensible. _____

c. Would have to guess often. _____

d. Verbal information essential. _____

NOTE: Verbal Information is that which is presented by voice or captions and include words embedded in pictures or "written" on

III. VERBAL CHARACTERISTICS

14. Verbal information load. (Check one):

High _____ Medium _____ Low _____

15. English skills required (Check one):

High _____ (College) Medium _____ (High school) Low _____ (Primary)

16. Technical vocabulary assumed of viewer (Check one):

Adv. _____ Inter. _____ Beg. _____

NOTE: Terms used that are not defined.

17. If Captioned: Does screen positioning of captions show which actor is speaking? (If film has only one speaker, check "Yes") Check one:

Yes _____ Sometimes _____ No _____

18. If Captioned: Are captions of new concepts and vocabulary distinguishable from rest of text? (Check one):

Distinct _____ (all caps; underlined; or out of text)

Marginally Distinct _____ (Cap 1st letter or has slightly different font)

Not Distinct _____ (Close to or same as text)

19. If Captioned: Consistency of ways used to caption new concepts/vocabulary (Check one):

Consistent _____ Moderately Consistent _____ Inconsistent _____

20. If Captioned: Does pacing of shots or captions interfere with reading of captions? (Applies to rate of cutting, action within shots, or busy captions) Check one:

Yes _____ Sometimes _____ No _____

IV. ORGANIZATIONAL CHARACTERISTICS

	Clearly	Sort of Clearly	Unclearly
21. Are the purposes of the presentation introduced clearly in the beginning?	1	2	3
22. Are summaries given when appropriate in the presentation?	1	2	3
23. Is the presentation well organized?	1	2	3

Additional Comments: